POSITION, ROTATION, AND SCALE INVARIANT RECOGNITION OF IMAGES USING HIGHER-ORDER SPECTRA

Vinod Chandran and Stephen L. Elgar

Washington State University, Pullman, WA 99164-2752 USA

ABSTRACT

A new approach to recognition of images using invariant features based on higher-order spectra is presented. Higher-order spectra are translation invariant because translation produces linear phase shifts which cancel. Scale and amplification invariance are satisfied by the phase of the integral of a higher-order spectrum along a radial line in higher-order frequency space because the contour of integration maps onto itself and both the real and imaginary parts are affected equally by the transformation. Rotation invariance is introduced by deriving invariants from the Radon transform of the image and using the cyclic-shift invariance property of the discrete Fourier transform magnitude. Results on synthetic and actual images show isolated, compact clusters in feature space and high classification accuracies.

1. INTRODUCTION

Efficient extraction of global invariant features will facilitate real-time classification of objects in an uncluttered environment. Higher-order correlations and spectra [1] offer a rich source of information about the shape of an object to extract such features from. Higher-order correlation and spectral techniques have been proposed for object detection and classification [2-5]. These techniques rely on higher-order correlation energies and outputs of normalized matched filters and do not exploit the information in the structure of the higher-order correlation function to the extent possible. The algorithm presented in this study extracts invariant features using such information. The algorithm is explained in section 2. Sections 3 and 4 describe applications to synthetic and actual images, respectively, followed by a conclusion in section 5.

2. THE ALGORITHM

Let $g(u,v)$ be an NxN image. Let $\{x_\theta(n)\}$ be the Radon transform of the image, *i.e.*, the set of parallel beam projections of the image on to lines at angle θ with respect to the horizontal axis. Assume that the projections are computed at equal increments of angle θ, *i.e.*, $\theta_i = \frac{i\pi}{N_\theta}$ for $i = 0$ to $N_\theta - 1$. The image is thus reduced to a set of one dimensional functions.

2.1 Invariance Properties

It can be readily inferred that

(a) a shift in the 2D image results in a shift in every projection except for one parallel to the direction of the shift,
(b) a scale change of the 2D image in the direction of projection results in multiplication of the 1D projection by a constant,
(c) a scale change of the 2D image perpendicular to the direction of projection results in a scale change of the 1D projection, and
(d) a rotation of the 2D image results in a cyclic shift in the set of projection functions.

Therefore, features $P(a)$ that are invariant to shift, scaling or amplification of the 1D functions $x(n)$ will provide a set $\{P(a)(\theta), a = 1, Na\}$ of 1D functions of θ, which are cyclically shifted in the interval $[0, \pi)$ when the image is rotated.

2.2 Bispectral Invariants for 1D Sequences

The *bispectrum* $B(f_1, f_2)$ of a one-dimensional, deterministic, discrete-time signal, $x(n)$, is defined as

$$B(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2) \qquad (1)$$

where $X(f)$ is the discrete-time Fourier transform of $x(n)$ and f is frequency normalized by one half of the sampling frequency. By virtue of its symmetry properties, the bispectrum of a real signal is unique in the triangular region of computation, $0 \le f_2 \le f_1 \le f_1 + f_2 \le 1$, provided there is no bispectral aliasing.

Parameters $P(a)$, that are translation, DC level, amplification, and scale invariant are defined from the bispectrum of $x(n)$ as follows :

$$P(a) = \arctan\left(\frac{I_i(a)}{I_r(a)}\right) \qquad (2)$$

where

$$I(a) = I_r(a) + j\, I_i(a) = \int_{f_1=0^+}^{1/(1+a)} B(f_1, af_1)\, df_1 \qquad (3)$$

for $0 < a \le 1$, and $j = \sqrt{-1}$.

Note that the bispectral values are integrated along straight lines with slope a passing through the origin in the bifrequency space. It has been shown [6,7] that $P(a)$ are invariant to translation, amplification or DC-level change in $x(n)$ unconditionally and invariant to scale change in $x(n)$ provided $x(n)$ is bandlimited and the scaling does not inroduce any aliasing.

0-7803-0532-9/92 $3.00

A scale change in the time domain shrinks or stretches the bispectrum equally in f_1 and f_2 and the line $f_2 = af_1$ maps on to itself. The integral of the bispectrum along this line is therefore merely multiplied by a real constant leaving its phase unchanged, in the absence of aliasing. It is interesting to consider this operation in the time domain. The bispectrum along a radial line in bifrequency space corresponds to the Fourier transform of a parallel beam projection of the triple correlation function. The real part corresponds to the even component of the projection function and the imaginary part to the odd component. $P(a)$ are thus ratios of the L_1 norms (or sum of amplitudes) of the spectra of the odd component and the even component of projections of the triple correlation function. Since scaling affects odd and even components equally, this ratio is unchanged. Thus, $P(a)$ extract information from the structure of the triple correlation function and serve as better features for classification than mere energy in the triple correlation function.

If the input is not strictly bandlimited as assumed above, the parameters $P(a)$ will not be strictly invariant for scaled inputs. This variation can be accounted for by (a) taking this variation into consideration during the training phase of the classifier, or (b) transforming the input sequence $x(n)$ into a new sequence $y(n)$ which is bandlimited, without losing the invariance properties.

This transformation need not be unique provided it is unique between the classes. One such transformation of the input sequence $x(n)$ is achieved by discarding the phases and zero padding the magnitude of the Fourier transform $\mid X(k) \mid$ of $x(n)$ for positive frequencies only, to yield a sequence $y(n)$. The parameters $P(a)$ computed from $y(n)$ satisfy the same invariance as those computed from $x(n)$. These parameters may now be used as features to discriminate between the original sequences.

Although, uniqueness is lost in the transformation owing to the loss of phase information, features derived from $y(n)$ can still be used for classification, because
(a) The features are sensitive enough to detect even small changes in the shape of $x(n)$.
(b) Images are classified from a set of projections, not a single projection.
(c) The integration of the bispectrum in the bifrequency plane provides some averaging, and the expected value of the bispectrum for white Gaussian noise is zero. Therefore, the features are insensitive to uncorrelated additive Gaussian noise.
(d) A least mean square error approximation to $x(n)$, or a periodic version of it as demanded by the discrete Fourier transform, in the class of harmonic functions will yield essentially the same $y(n)$ except for the truncation of the Fourier series to a smaller number of terms. If the spectrum has fallen considerably at the point of truncation, $y(n)$ will not differ significantly in shape, and therefore the features will not deviate significantly from their true values. In other words, the features extracted from least mean square error approximations to a given $x(n)$ will form a cluster in the feature space.

The computational procedure to calculate $P(a)$ given an N-point real sequence $x(n)$ is described in [6]. It involves approximation of the integral in equation 3 by a sum and bilinear interpolation of the bispectrum.

2.3 Invariant Features from Images

The Fourier Slice theorem [8] states that the Fourier transform of the projection of an image on to a line at angle θ to the x-axis is the 2D Fourier transform of the image itself evaluated along a radial line (*i.e.*, θ constant). Therefore, instead of using the projection sequences themselves their discrete Fourier transform magnitudes excluding the DC component are used to derive the invariant features. This has three advantages, *viz.*, (a) raster rotations to get the projection functions are eliminated, (b) the features are also invariant to a constant change in the background of the image, and (c) scale invariance is better satisfied. An N x N image $g(u, v)$ is Fourier transformed using a 2D FFT routine. The 2D Fourier transform is then transformed to a polar grid using bilinear interpolation, yielding $G(r, \theta)$, $r = 0$ to $N/2$ and $\theta = \frac{\pi i}{N_\theta}$ for $i = 0$ to $2N_\theta - 1$. The Fourier transform magnitude along a radial line is zero padded to form the sequence

$$y_\theta(n) = \begin{cases} |G(n+1, \theta)|, & \text{for } n = 0, 1, ..N/2 - 1; \\ 0, & \text{for } N/2 <= n < N. \end{cases}$$

Bispectral invariants are computed for this sequence to yield $P_\theta(a)$, $a = 1, 2, ..N_a$. This procedure is repeated for different angles, $\theta = \frac{i\pi}{N_\theta}$ for $i = 0$ to $N_\theta - 1$. In order to obtain rotation invariance, these parameters are then considered as one dimensional sequences for each fixed value of parameter a. An N_θ-point DFT is computed for each such sequence and the magnitude of the Fourier transform for positive frequencies, including the DC value, is zero padded to N_θ points. Bispectral invariants are then computed for these sequences, yielding the set of features, $P(a)(b)$, for $a = 1, .., N_a$ and $b = 1, .., N_b$. These features are translation and scale invariant owing to the properties described in section 2.1 and rotation invariant owing to the cyclic shift invariance of the DFT.

In practice, the truncation of the discrete-time Fourier transform and aliasing involved in the computation of the DFT result in some variation of the features with scale. It has been shown [9] that essential information is retained in the discrete Fourier transform despite aliasing provided the size of the object and its details of interest lie between about 4 points and $N/4$ points. The feasibility of this algorithm was tested with both synthetic and real data.

3. APPLICATION TO SYNTHETIC IMAGES

The input consisted of 128x128 bi-level images, with 0 value in the background and a constant nonzero value within the shape of interest. Four regular geometric shapes were chosen for classification, a square, a circle, a plus sign, and the alphabet 'H', as shown by the prototype images in Figure 1. Randomly translated and rotated versions of the prototype images were generated. The translation was uniform random in [-8,8] and the rotation was uniform random in $[0, \pi)$. 256 projections were considered and each image was

reduced to an 8x8 set of invariant features. Figures 2 and 3 show the values of three of these features, P(1)(1), P(4)(8), and P(8)(8), for the inputs from Figure 1 at scalefactors 1 and 0.5 respectively.

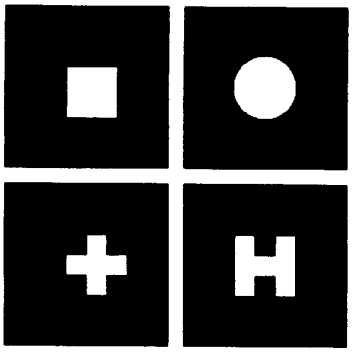

Figure 1. Prototype Synthetic Images

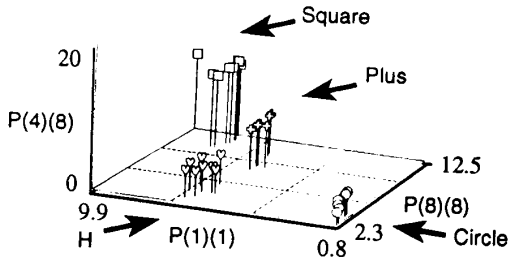

Figure 2. Features for random rotated and translated versions of the images in Figure 1

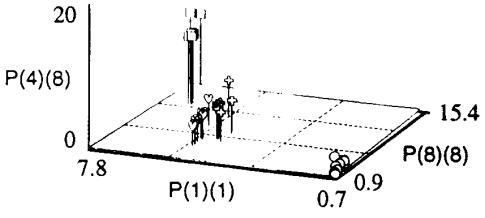

Figure 3. Features for random rotated and translated versions of the images in Figure 1 at scalefactor 0.5

There were 8 randomly translated and rotated features from each pattern. Note that the features form well defined clusters in the feature space. Gaussian noise was then added to the random realizations. Figures 4 and 5 show the dispersion of the clusters of features as the SNR falls to 40db and 20dB respectively. A 3NN classfier [10,11] was trained using the 20dB SNR data set and tested using inputs with various SNR's. The classification accuracy results are presented in Table 1.

Table 1
Classification Accuracy (20dB SNR Training)

SNR	Misclassifications	Accuracy
∞dB	0/32	100%
40dB	0/32	100%
20dB	1/64	98%
10dB	13/32	60%

It is possible to achieve high classification accuracies for even very low SNR inputs by averaging the 2D spectrum before feature extraction or averaging the bispectra, thereby increasing the effective SNR to better than 20dB. To facilitate this, a high resolution input may be split into an ensemble of lower resolution images.

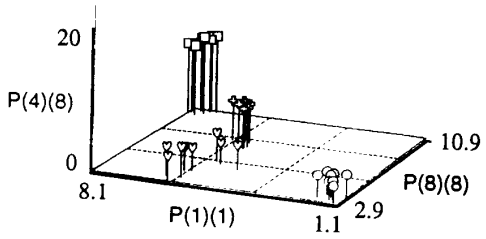

Figure 4. Features for random rotated and translated versions of the images in Figure 1 with SNR 40dB

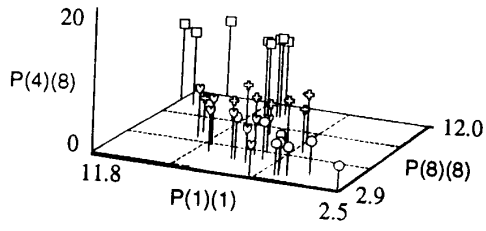

Figure 5. Features for random rotated and translated versions of the images in Figure 1 with SNR 20dB

4. APPLICATION TO ACTUAL IMAGES

The input to this test consisted of 256x256 images acquired using a video camera and digitizer. The intensity of each pixel was quantized to 256 values or 1 byte. There were three classes of input : a square nut, a hexagonal nut and a circular washer, as shown in Figure 6. The images acquired were noisy owing to nonuniform illumination and nonuniform reflectance of the surfaces involved and had a texture in the background. No preprocessing was done to either remove noise or to enhance edges. Each class was imaged at three different scale factors, in the ratio 0.7 : 0.9 : 1.1, depending on the distance of the video camera from the object. 12 images of each class (four for each scale factor) were acquired, and additional images were generated by random translation in [-16,16] and random rotation in $[0, \pi)$ to yield 48 images from each class. 256 projections were considered and an 8x8 set of invariant features was extracted from each image. A Principal component analysis [12] was performed to reduce the dimensionality of the feature vector. A scatter plot of the principal components of the features extracted from the 48 images of each of the three classes is shown in Figure 7. Isolated clusters are again formed for each class, indicating that 100% classification accuracy is possible.

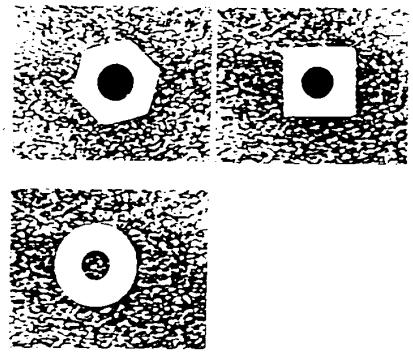

Figure 6. Prototype Actual Images

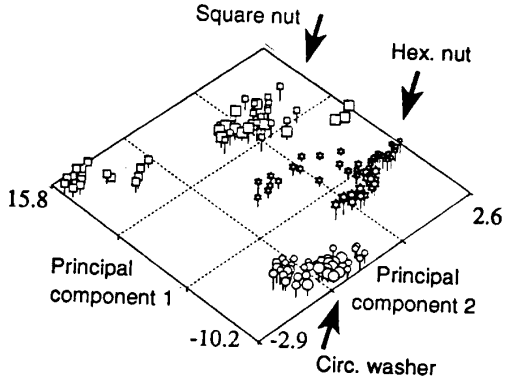

Figure 7. Principal components of the features for random rotated and translated versions of the images in Figure 6

5. CONCLUSION

A new algorithm for extracting translation, rotation, and scale invariant features from images is presented. Test results indicate the potential for high classification accuracies with standard classifiers even for degraded inputs. A complexity analysis of the algorithm and a parallel implementation can be found in [13]. Although the algorithm is presented using bispectral invariant features, it is easily generalized to extract higher order spectral invariants.

REFERENCES

1. C.L. Nikias and M.R. Raghuveer, "Bispectrum Estimation: A Digital Signal Processing Framework," *Proceedings of the IEEE*, 75, pp. 869-889, 1987.
2. L. Capodiferro, R. Gusani, G. Jacoviti and M. Vascotto, "A correlation based technique for shift, scale and rotation independent object identification," *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 221-224, Dallas, TX, 1987.
3. M.K. Tsatsanis and G.B. Giannakis, "Translation, rotation and scaling invariant object and texture classification using polyspectra," *Proc. of SPIE'90* , 1348, pp. 103-115, San Diego, CA, July 1990.
4. G.B. Giannakis and M.K. Tsatsanis, "Signal Detection using matched filtering and higher-order statistics," *IEEE Transactions on ASSP*, 38, pp. 1284-1296, July 1990.
5. M.J. Hinich, "Detecting a transient signal by bispectral analysis," *IEEE Transactions on ASSP*, 38, pp. 1277-1283, July 1990.
6. Vinod Chandran and Steve Elgar, "Shape Discrimination Using Invariant Features Defined From Higher Order Spectra," *Proc. of ICASSP'91*, Toronto, vol. 5, pp. 3105-3109, May 14-18, 1991.
7. Vinod Chandran and Steve Elgar, "Pattern Recognition Using Invariants Defined From Higher Order Spectra - One dimensional Inputs," *IEEE Transactions on Signal Processing*, accepted for publication.
8. Azriel Rosenfeld and A.C. Kak, *Digital Picture Processing* vol. 1, pp.353-369, Academic Press, 1982.
9. Jurgen Altmann and Herbert J.P. Reitbock, "A Fast Correlation Method for Scale- and Translation-Invariant Pattern Recognition," *IEEE Transactions on PAMI*, vol. PAMI-6, January 1984.
10. R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley Interscience, 1973.
11. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd Ed., Academic Press, 1990.
12. C.R. Rao, "The Use and Interpretation of Principal Component Analysis," *Sankhya A*, vol. 26, pp. 329-358, 1964.
13. Vinod Chandran, *Affine Invariant Object Recognition - Algorithm and Parallel Implementation*, MSCS Project Report, W.S.U., Pullman, WA, 1991.